Xuxu Pan

408.421.5402 · x_pan0427@hotmail.com · San Francisco, CA

EDUCATION

M.S. in Data Science, University of San Francisco — June, 2020 (expected)
Studied Machine Learning, Deep Learning, Distributed Computing (Pyspark), Data Structures and Algorithms, SQL, Python, A/B test, Time Series — San Francisco, USA

M.S. in Accounting and Finance, London School of Economics — July, 2014
Studied Stochastic Process, Financial Modeling — London, UK

B.A. in Accounting and Journalism, Sun Yat-sen University — June, 2013.
Studied Advanced Mathematics (Calculus, Linear Algebra, Statistics) — Guangzhou, China

WORK EXPERIENCE

Propeller Health — Dec, 2019 - now
Data Science Intern — San Francisco, USA

- Developed gradient boosted models to prioritize customer support tickets through NLP on text data from SQL queries to tables with over 1.5 billion rows on Amazon Redshift with Python and PyTorch (work in progress).
- Published several live dashboards to Tableau Server to better inform the customer experience team about how patients are using the customer support platform (work in progress).

E Fund Management Co. Ltd. — Mar, 2015 - Jul, 2019
Fund Accountant — Guangzhou, China

- Built automated data pipelines from external data sources including Bloomberg and Reuters to the portfolio accounting system with SQL. Analyzed trading, security master, price and corporate action data from those external data sources.
- Designed workflow and key features of the systems, tested the features with software engineer colleagues and system vendors, improved operation and portfolio accounting efficiency by 30%.

PROJECTS

- **SARIMA Model on California House Price Prediction**: Predicted median sold price of house in California with a SARIMA model using Statsmodels in Python, and obtained an RMSE of 2,511.58 on the validation set.
- **XGBoost Model on Heart Rate and Blood Pressure Prediction**: Predict mean measurement of heart rate and mean blood pressure of critically ill patients from time-series features with XGBoost model using Scikit-learn in Python, and obtained an R-squared around 0.93 on the validation set.
- **H2ODeepLearning Model on Air Quality Index Prediction**: Predicted Air Quality Index (AQI) from time-series features with H2ODeepLearning estimator model using Spark ML and H2O Sparkling water, and obtained a median error around 12 (300 based) on the validation set.

SKILLS

Python, Pytorch, SQL, Pandas, Pyspark, Tableau**,** Excel with VBA and DEVELOPER

CERTIFICATE

Passed all three levels of Chartered Financial Analyst (CFA) exams.